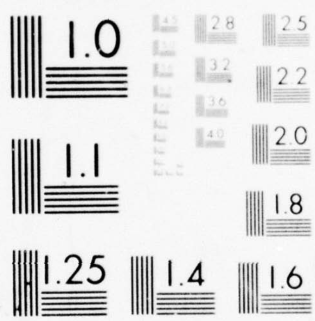
AD-A039 022 SYRACUSE UNIV N Y DEPT OF INDUSTRIAL ENGINEERING AND--ETC F/G 9/2
SUMMARY OF TECHNICAL PROGRESS ON BAYESIAN SOFTWARE PREDICTION M--ETC(U)
MAR 77 A L GOEL F30602-76-C-0097
UNCLASSIFIED TR-76-18 RADC-TR-77-112 NL

1 OF 1
AD A039022

END
DATE FILMED
5-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

12
NW

RADC-TR-77-112
Technical Report
March 1977

SUMMARY OF TECHNICAL PROGRESS ON BAYESIAN SOFTWARE PREDICTION MODELS

Syracuse University

ROME AIR DEVELOPMENT CENTER
AIR FORCE SYSTEMS COMMAND
GRIFFISS AIR FORCE BASE, NEW YORK 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.

This report has been reviewed and is approved for publication.

APPROVED: Alan N. Sukert

ALAN N. SUKERT, Captain, USAF
Project Engineer

APPROVED: Alan R Barnum

ALAN R. BARNUM
Assistant Chief
Information Sciences Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-77-112	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SUMMARY OF TECHNICAL PROGRESS ON BAYESIAN SOFTWARE PREDICTION MODELS		5. TYPE OF REPORT & PERIOD COVERED Interim Report 8 Dec 75 - 8 Dec 76
		6. PERFORMING ORG. REPORT NUMBER 76-18
7. AUTHOR(s) Amrit L. Goel (see reverse)		8. CONTRACT OR GRANT NUMBER(s) F30602-76-C-0097
9. PERFORMING ORGANIZATION NAME AND ADDRESS Syracuse University Syracuse NY 13210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62702F 55811403
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (ISIS) Griffiss AFB NY 13441		12. REPORT DATE March 1977
		13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Captain Alan N. Sukert (ISIS)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian Models
Software Correction Limit Policies
Software Models
Debugging Models
Imperfect Debugging
Software Reliability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Work has been completed on development of Bayesian Software Correction Limit Policies designed to determine the optimum time value that minimizes the long run average cost of debugging at two levels - correction action undertaken by the programmer (Phase I) and action undertaken by a system analyst or system designer, if the error is not corrected in Phase I (Phase II). Two models are developed - one assuming the cost of observations of error occurrence and next page

409 184

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

cont → correction time, prior to implementation of the optimum policy, is negligible; the other incorporating the cost of observations.

Work was also completed on an Imperfect Software Debugging Model that assumes errors are not corrected with certainty. By assuming the initial number of errors, probability of successfully correcting an error, and constant error occurence rate are all known, formulas for such quantities as distribution of time to completely debugged software, distribution of time to a specified number of remaining errors, and expected number of errors detected by time t can be derived.

Work is currently in progress in extending the Imperfect Debugging Model to incorporate error correction time, estimation of model parameters and development of a Bayesian model; developing bivariate software reliability models where system errors are classified as serious and non-serious; development of empirical models for software error data; development of software reliability demonstration plans for making accept/reject decisions for software packages; and investigating the effects of changes in prior distributions and/or model parameters on quantities of interest.

Block 7: Associate Professor, Department of Industrial Engineering and Operations Research; and School of Computers and Information Science, Syracuse University, Syracuse, New York 13210.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This interim report provides a summary of the technical activities pursued under contract F30602-76-C-0097 with RADC during January-December 1976. Sections 2 and 3 contain a brief description of the work on Bayesian software correction limit policies and a software reliability model with imperfect debugging, respectively. Section 4 lists the current work and plans for the next period.

2. BAYESIAN SOFTWARE CORRECTION LIMIT POLICIES

This study deals with the problem of determining an optimum correction limit policy for a large software system which is subject to random occurrences of errors. When an error occurs, a corrective action is undertaken to remove it. Such an action can be scheduled at two levels, which we call Phase I and Phase II. By Phase I we mean that the corrective action will be undertaken by the programmer while Phase II action is undertaken by a system analyst or system designer. First, Phase I corrective action is scheduled for a specified time T. If the error is not corrected in this time, it is referred to Phase II. Figure 2.1 shows the sequence of corrective actions in an operational phase. The objective is to determine the optimum value T^* of T which minimizes the long run average cost. Two models are developed for this purpose. In the first model (Model 1) we assume that the cost of observations of error occurrence and correction time, prior to the implementation of the optimum policy, is negligible. The second model (Model 2) incorporates the cost of observations.

In the following sections we provide a brief description of the various aspects of this problem. A detailed discussion will be given in a forthcoming technical report.

2.1 Model Development: Negligible Cost of Observations (Model 1)

2.1.1 Assumptions

In the development of the model, the following assumptions are made.

(i) The error occurrence time in a software system has an exponential distribution with an unknown mean λ.

(ii) The error correction time at Phase I is exponential with an unknown mean μ_1.

(iii) The Phase II error correction time has a general distribution with a known mean μ_2.

(iv) Reasonable prior distributions can be chosen for λ and μ_1.

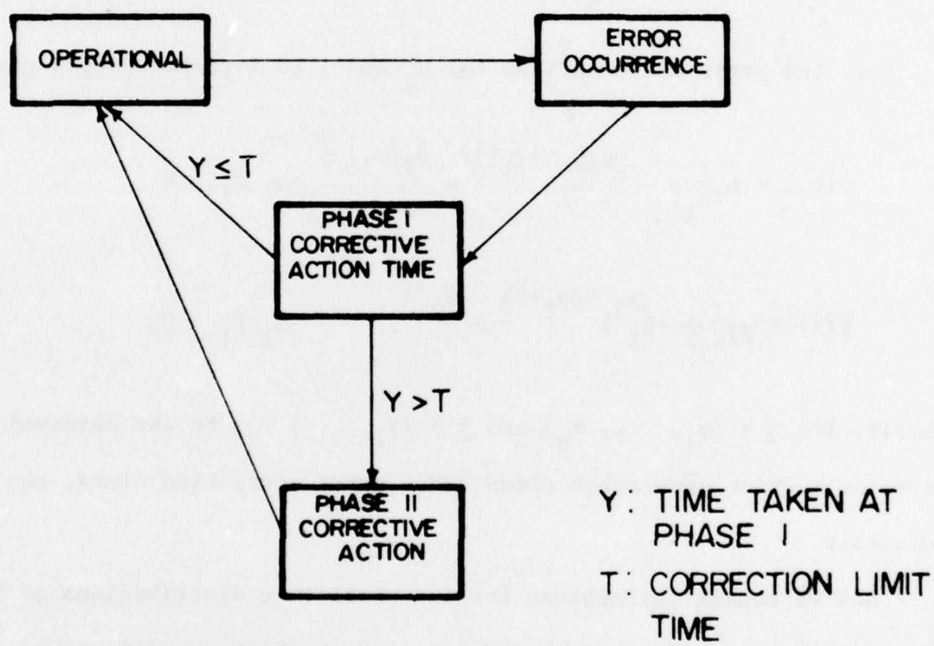


Figure 2.1. Sequence of Corrective Actions in Operational Phase.

2.1.2 Predictive Distributions

Let the random variables X and Y denote the error occurrence time and the Phase I error correction time, respectively, with probability density functions

$$f(x|\lambda) = \frac{1}{\lambda} \cdot e^{-x/\lambda}, \qquad x > 0,\ \lambda > 0$$

and

$$g(y|\mu_1) = \frac{1}{\mu_1} e^{-y/\mu_1}, \qquad y > 0,\ \mu_1 > 0.$$

Let the prior distributions for μ_1 and λ be inverted gammas given by

$$p(\mu_1) = \frac{1}{\Gamma(\alpha_1)} \cdot \beta_1^{\alpha_1} \mu_1^{-(\alpha_1+1)} e^{-\beta_1/\mu_1}, \qquad \alpha_1, \beta_1 > 0$$

and

$$p(\lambda) = \frac{1}{\Gamma(\alpha_2)} \cdot \beta_2^{\alpha_2} \lambda^{-(\alpha_2+1)} \cdot e^{-\beta_2/\lambda}, \qquad \alpha_2, \beta_2 > 0.$$

Finally, let $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$ be the observed values of n error occurrence times and n error correction times, respectively.

Now we obtain expressions for the predictive distributions of Y and X and the Bayesian estimates $\hat{y}_{n+1}$ and $\hat{x}_{n+1}$ which we will use to obtain the cost function.

For given observations $\underline{y}$, the likelihood function of μ_1 is

$$\ell(\mu_1|\underline{y}) \propto \mu_1^{-n} \cdot e^{-\sum_{i=1}^{n} y_i/\mu_1},$$

and the posterior distribution of μ_1 is

$$p(\mu_1|\underline{y}) = \frac{(\beta_1 + \sum_{i=1}^{n} y_i)^{\alpha_1+n}}{\Gamma(\alpha_1+n)} \cdot \mu_1^{-(\alpha_1+n+1)} \, e^{-\{(\beta_1+\sum_{i=1}^{n} y_i)/(\mu_1)\}}$$

Then, the predictive distribution of error correction time at Phase I is

$$g(y|\underline{y}) = \left(\frac{n + \alpha_1}{\sum_{i=1}^{n} y_i + \beta_1}\right)\left(1 + \frac{y}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-(n+\alpha_1+1)}$$

and the cumulative predictive distribution to some specified time t is

$$G(t|\underline{y}) = \int_0^t g(y|\underline{y})dt = 1 - \left(1 + \frac{t}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-(n+\alpha_1)}$$

We define the predictive Phase I error correction rate as

$$r(t|\underline{y}) = \frac{g(t|\underline{y})}{\bar{G}(t|y)},$$

so that

$$r(t|\underline{y}) = \left(\frac{n + \alpha_1}{\sum_{i=1}^{n} y_i + \beta_1}\right)\left(1 + \frac{t}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-1}$$

where

$$\bar{G}(t) = 1 - G(t)$$

From the above results the Bayesian estimate of the (n+1)st error correction time for given $\underline{y}$ is

$$\hat{y}_{n+1} = \frac{\sum_{i=1}^{n} y_i + \beta_1}{\alpha_1 + n - 1}$$

Proceeding similarly, the Bayesian estimate of the time to (n+1)st error occurrence, for given $\underline{x}$, is

$$\hat{x}_{n+1} = \frac{\sum_{i=1}^{n} x_i + \beta_2}{\alpha_2 + n - 1}$$

2.1.3 Cost Function

Let $c_1(c_2)$ be the cost per unit time of error correction in Phase I (Phase II) and the costs be linear functions of time. If we consider one cycle to be the time from the beginning of (n+1)st operation to the beginning of (n+2)nd operation, then the expected cost in one cycle is

$$E(C) = c_1 \int_0^T \bar{G}(t|\underline{y})dt + c_2\, \mu_2\, \bar{G}\,(T|\underline{y}),$$

where T denotes the scheduled correction limit time in Phase I.

The expected length of one cycle is

$$E(T) = \hat{x}_{n+1} + \int_0^T \bar{G}(t|\underline{y})dt + \mu_2\, \bar{G}(T|\underline{y})\ ,$$

and hence the long run expected cost per unit time is

$$C(T) \equiv \frac{E(C)}{E(T)}$$

or

$$C(T) = \frac{c_1 \int_0^T \bar{G}(t|\underline{y})dt + c_2\, \mu_2\, \bar{G}(T|\underline{y})}{\hat{x}_{n+1} + \int_0^T \bar{G}(t|\underline{y})dt + \mu_2\, \bar{G}(T|\underline{y})}$$

2.1.4 Optimum Policy

From the above expressions, we note that

$$C(0) = \frac{c_2\, \mu_2}{x_{n+1} + \mu_2}$$

and

$$C(\infty) = \frac{c_1\, \hat{y}_{n+1}}{\hat{x}_{n+1} + \hat{y}_{n+1}}$$

where $\hat{y}_{n+1}$ is the Bayesian estimate of **y** for given data $\underline{y}$. Also, note that $T = 0$ means that the errors are corrected only at Phase II while $T = \infty$ means that they are corrected at Phase I.

To obtain an optimum T* which minimizes the long run average cost per unit time, C(T), we need the following theorems and corollary. They are given here without proofs.

Theorem 2.1:

Assume $c_1 < c_2$. Then there exists a finite and unique T* which satisfies

$$r(T|\underline{y})\ \{c_2\ \hat{x}_{n+1} + (c_2-c_1) \int_0^T \bar{G}(t|\underline{y})dt\} + (c_2-c_1)\ \bar{G}(T|\underline{y}) = \frac{c_1\ \hat{x}_{n+1}}{\mu_2}$$

under the following condition

$$r(0|\underline{y}) > \frac{c_1(\hat{x}_{n+1} + \mu_2) - c_2\ \mu_2}{c_2\ \hat{x}_{n+1}\ \mu_2}$$

Theorem 2.2:

If the above conditions are satisfied then there also exists a finite and unique upper bound $\bar{T}(>T^*)$ such that

$$r(\bar{T}|\underline{y}) = \frac{c_1\ \hat{x}_{n+1}}{\mu_2\{c_2\ \hat{x}_{n+1} + (c_2-c_1)\mu_1\}}$$

This upper bound can be used to obtain an initial value for solving the nonlinear equations in T*.

Corollary 2.3

If there exists an optimum T* then the associated cost function is given by

$$C(T^*) = \frac{c_1 - c_2\ \mu_2\ r(T^*|\underline{y})}{1 - \mu_2\ r(T^*|\underline{y})}$$

2.1.5 Numerical Example

We use simulated data in this example to illustrate the calculation and nature of various quantities in the determination of T*.

Let

$c_1 = 8000 \qquad c_2 = 9000$

$\alpha_1 = 0 \qquad \beta_1 = 0$

$\alpha_2 = 0 \qquad \beta_2 = 0$

$\mu_2 = 0.7$

The simulated data (x_n, y_n) are given in Table 2.1. Suppose n = 10 data points are available. The Bayesian estimates of failure time and correction time are $\hat{x}_{11} = 59.60$ and $\hat{y}_{11} = 0.78$, respectively. Such values for various n are given in Table 2.2 . For the case n = 10 we see that the optimum correction limit time is T* = 0.90 hours and the corresponding minimum cost rate is C(T*) = 99.44 dollars/hour.

Thus, for this set of data, we will schedule corrective action in Phase I for 0.90 hours and if it cannot be completed in this time, the software system will be referred to the system programmer for corrective action.

Table 2.1

Simulated Values of x_n and y_n

n	x_n (Hrs.)	y_n (Hrs.)	n	x_n (Hrs.)	y_n (Hrs.)
1	61.34	1.90	11	53.44	1.03
2	27.84	1.08	12	2.87	0.95
3	154.30	0.85	13	31.27	0.60
4	14.58	0.26	14	97.06	0.02
5	10.86	0.01	15	78.17	1.49
6	35.35	0.31	16	124.52	0.52
7	140.13	0.38	17	0.49	0.36
8	36.47	1.50	18	12.33	0.08
9	8.74	0.43	19	85.44	3.51
10	46.79	0.27	20	23.59	0.10

Table 2.2

Calculation for the Optimum Correction Time Policy

n	$\hat{x}_{n+1}$ (hr.)	$\hat{y}_{n+1}$ (hr.)	T* (hr.)	C(T*)
2	89.17	2.98	0	70.10
3	121.74	1.92	0	51.45
4	86.02	1.36	0	72.65
5	67.23	1.02	0	92.74
6	60.85	0.88	0.32	101.05
7	74.07	0.80	0.73	80.11
8	68.70	0.90	0.02	90.78
9	61.20	0.84	0.38	100.60
10	59.60	0.78	0.90	99.44
11	58.98	0.80	0.66	102.87
12	53.88	0.82	0.50	113.82
13	52.00	0.80	0.68	116.85
14	55.46	0.74	1.45	103.80
15	57.09	0.79	0.75	106.51
16	61.58	0.77	1.02	97.47
17	57.76	0.75	1.45	101.22
18	55.09	0.71	2.16	101.14
19	56.78	0.86	0	109.61
20	55.03	0.82	0.13	112.97

2.2. MODEL DEVELOPMENT INCLUDING SAMPLING COST (MODEL 2)

In this section we develop the cost model for the optimum correction limit policy by incorporating sampling cost. It is assumed that, in addition to the assumptions of Section 2.1.1, the sampling cost is a linear function of sample size.

Let c be the sampling cost at each stage, and $C_n(T_n)$ be the expected cost per unit time until the completion of (n+1)st corrective action under scheduling time T_n. If we decide to take another observation ((n+1)st), then $C_{n+1}(T_{n+1})$ is the cost rate function until the completion of (n+2)nd corrective action under scheduling time T_{n+1}.

2.2.1 Cost Function

Following a procedure similar to that of Sections 2.1.2 and 2.1.3, the expected cost per unit time at the end of the (n+1)st corrective action is

$$C_n(T_n) = \frac{nc + c_1 \int_0^{T_n} \bar{G}(t|\underline{y})dt + c_2\mu_2 \bar{G}(T_n|\underline{y})}{\sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i + \int_0^{T_n} \bar{G}(t|\underline{y})dt + \mu_2 \bar{G}(T_n|\underline{y})}$$

If we decide to take the next, (n+1)st, observation, then the cost rate function to the end of (n+2)nd corrective action is similarly obtained as

$$C_{n+1}(T_{n+1}) = \frac{(n+1)c+c_1 \int_0^{T_{n+1}} \bar{G}(t|\underline{y})dt + c_2 \mu_2 \bar{G}(T|\underline{y})}{\sum_{i=1}^{n} x_i + 2\hat{x}_{n+1} + \sum_{i=1}^{n} y_i + \hat{y}_{n+1} + \int_0^{T_{n+1}} \bar{G}(t|\underline{y})dt + \mu_2 \bar{G}(T_{n+1}|\underline{y})}$$

2.2.2 Optimum Policy

Our objective is to determine the optimum sample size n^* and the optimum correction limit T_n^* such that

$$C_n(T_n^*) \leq C_{n+1}(T_{n+1}^*)$$

Given that n observations have been taken, the following steps summarize the procedure of determining these quantities:

(i) Calculate $C_n(T_n^*)$ and $C_{n+1}(T_{n+1}^*)$

(ii) If $C_n(T_n^*) \leq C_{n+1}(T_{n+1}^*)$, then stop taking observations and employ n^* and T_n^* as the optimum policy.

(iii) If $C_n(T_n^*) > C_{n+1}(T_n^*)$, take (n+1)st observation, i.e. let $n = n+1$ and go to step (i).

The following theorems are useful in determining $C_n(T_n^*)$ and $C_{n+1}(T_{n+1}^*)$. These are given here without proofs.

Theorem 2.4

Suppose $c_1 < c_2$ and $A \geq 0$. Then there exists a unique and finite T_n^* satisfying

$$r(T_n|\underline{y}) \{A + (c_2-c_1) \int_0^{T_n} \bar{G}(t|\underline{y})dt\} + (c_2-c_1)\, \bar{G}(T_n|\underline{y}) = B$$

where

$$A = c_2 \{ \sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i\} - nc$$

$$B = \frac{1}{\mu_2}\left[c_1\{ \sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i\} - nc\right]$$

Also, the associated cost rate function is given by

$$C_n(T_n^*) = \frac{c_1 - c_2\, \mu_2\, r(T_n^*)}{1 - \mu_2\, r(T_n^*)}$$

Theorem 2.5

If the conditions of Theorem 2.4 hold, then there exists a finite and upper bound $\bar{T}_n$ ($> T_n^*$) such that

$$r(\bar{T}_n|\underline{y}) = \frac{B}{A + (c_2 - c_1)\hat{y}_{n+1}}$$

Theorem 2.6

If both T_n^* and T_{n+1}^* exist, then the following relationship holds:

$$C_n(T_n^*) \underset{(\leq)}{>} C_{n+1}(T_{n+1}^*)$$

$$\Longleftrightarrow \quad r(T_n^*) \underset{(\leq)}{>} r(T_{n+1}^*)$$

$$\Longleftrightarrow \quad T_n^* \underset{(\leq)}{>} T_{n+1}^*$$

2.2.3 Numerical Example

In this example we use simulated data to illustrate the determination of n^* and T_n^*. Simulated values of x_n and y_n for various n are given in Table 2.3. Let

$c_1 = 8000$, $c_2 = 9000$, $c = 40$

$\alpha_1 = .8$ $\beta_1 = 1$

$\alpha_2 = 0$ $\beta_2 = 0$

$\mu_2 = 0.7$

Then the calculated x_{n+1}, y_{n+1}, T_n^*, $C_n(T_n^*)$, T_{n+1}^* and $C_{n+1}(T_{n+1}^*)$ are computed from the above expressions and are given in Table 2.4. From this table we see that for n = 11, $C_{11}(T_{11}^*) = 21.74$ and $C_{12}(T_{12}^*) = 23.63$ so that

Table 2.3

Simulated Values of x_n and y_n

n	x_n(hr)	y_n(hr)	n	x_n(hr)	y_n(hr)
1	32.25	1.69	10	3.72	0.15
2	34.77	0.12	11	50.85	0.07
3	63.92	0.23	12	64.89	0.12
4	21.03	0.41	13	0.76	0.29
5	39.42	0.20	14	87.45	1.33
6	9.97	0.37	15	64.12	0.77
7	3.69	0.22	16	30.98	1.37
8	2.42	1.75	17	127.05	1.39
9	10.71	3.00	18	85.54	0.21

Table 2.4

Calculations for the Optimum Policy

n	$\hat{x}_{n+1}$	$\hat{y}_{n+1}$	T_n^*	$C_n(T_n^*)$	T_{n+1}^*	$C_{n+1}(T_{n+1}^*)$
2	67.01	1.56	0	46.73	0	37.72
3	65.47	1.09	0	32.24	0	32.05
4	50.65	0.91	0.33	30.9	0.33	27.06
5	47.85	0.76	0.92	24.44	0.92	23.74
6	40.27	0.69	0.12	23.00	0.13	20.55
7	34.17	0.62	0.19	21.43	0.19	19.12
8	29.64	0.77	0.95	25.63	0.95	23.25
9	27.27	1.02	0	26.21	0	24.93
10	24.66	0.93	0	26.23	0	24.32
11	27.27	0.85	0.09	21.74	0.09	23.63

$C_{11}(T^*_{11}) < C_{12}(T^*_{12})$. Therefore, the optimum policy is $n^* = 11$ and $T^*_n = 0.09$.

2.3 Concluding Remarks

In the previous sections we have given an overview of the key results for obtaining correction limit policies for two models. For the first model the sampling costs are assumed to be negligible while for the second, the sampling cost is taken to be a linear function of sample size. The main theorems have been given without proofs and the policy determination was illustrated via two numerical examples.

A detailed technical report on this task is being completed for submission to RADC.

3. A SOFTWARE RELIABILITY MODEL WITH IMPERFECT DEBUGGING

The purpose of this study is to develop software reliability models for the case when errors are not corrected with certainty. In other words, we are interested in developing and studying software reliability models for the case when the programmer sometimes fails to correct a detected error.

A description of the key results is given in the following subsections.

3.1 Assumptions

The following assumptions are made for developing the model.

(i) Errors in the software are independent of each other and have a constant occurrence rate λ.

(ii) The probability of 2 or more errors occurring at the same time is negligible. Then, if there are n errors in the software at present, the distribution of time to the next failure is

$$f(t) = n\lambda \cdot e^{-n\lambda t}$$

(iii) The correction time in the model is assumed to be zero. This assumption may not hold in some situations. We plan to develop another model which will include error correction time.

3.2 Model Development

Let p be the probability of successfully correcting an error,

q be the probability that the error is not corrected, $q = 1-p$,

$X(t)$ be a random variable denoting the system state,

N be the initial number of errors.

Let $P_{N,N-1}$ and $P_{N,N}$ represent the transition probabilities from state N to states N-1 and N, respectively. Then the cumulative distribution function (cdf) of the time to next error at state N is

$$F_N(t) = 1 - e^{-N\lambda t}$$

$$P_{N,N-1} = p$$

$$P_{N,N} = q$$

In general, the cdf of one step transition for the underlying Markov Process is:

$$Q_{ij}(t) = P_{ij}F_i(t),$$

where $\quad F_i(t) = 1 - e^{-i\lambda t},$

$$P_{ij} = \begin{cases} p & \text{if } j = i-1, \quad 1 \leq i \leq N \\ q & \text{if } j = i \\ 0 & \text{otherwise} \end{cases},$$

and

$$P_{0j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases}$$

This can be represented as:

$$(P_{ij}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ . \\ . \\ . \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & \cdots\cdots & N-1 & N \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & \cdots & & \\ p & q & 0 & & & \\ 0 & p & q & \ddots & & \\ & \ddots & \ddots & \ddots & & \\ & & & & & \\ & & & p & q & 0 \\ & & & 0 & p & q \end{array}\right] \end{array}$$

The above expressions constitute the basic model for the error process with parameters N, p and λ. For these expressions we obtain various quantities of interest in the following subsections.

3.3 Distribution of Time to a Completely Debugged Software

Let

$G_{i,0}(t)$ = cdf of first passage time from i to 0 by time t.

Then, by conditioning on the next one step, we get the renewal equations which we solve by Laplace-Stieltjes transforms. We get the cdf and pdf of first passage time as

$$G_{N,0}(t) = \sum_{j=1}^{N} C_j (1 - e^{-ip\lambda t})$$

and

$$g_{N,0}(t) = \sum_{j=1}^{N} C_j \, jp\lambda \, e^{-jp\lambda t},$$

$$\text{where } C_j = \binom{N}{i} (-1)^{j-1}$$

The cdf's of the first passage times for several values of p, N = 10, and $\lambda = 0.02$ are shown in Figure 3.1. As expected, the cdf for a larger p dominates that for a smaller p.

3.4 Distribution of Time to a Specified Number of Remaining Errors

In many situations we are not interested in a completely debugged program because of the cost involved. We may be willing to tolerate a certain number of remaining errors, say n_0, which will ensure some desired reliability. The distribution of time to n_0 is then of interest.

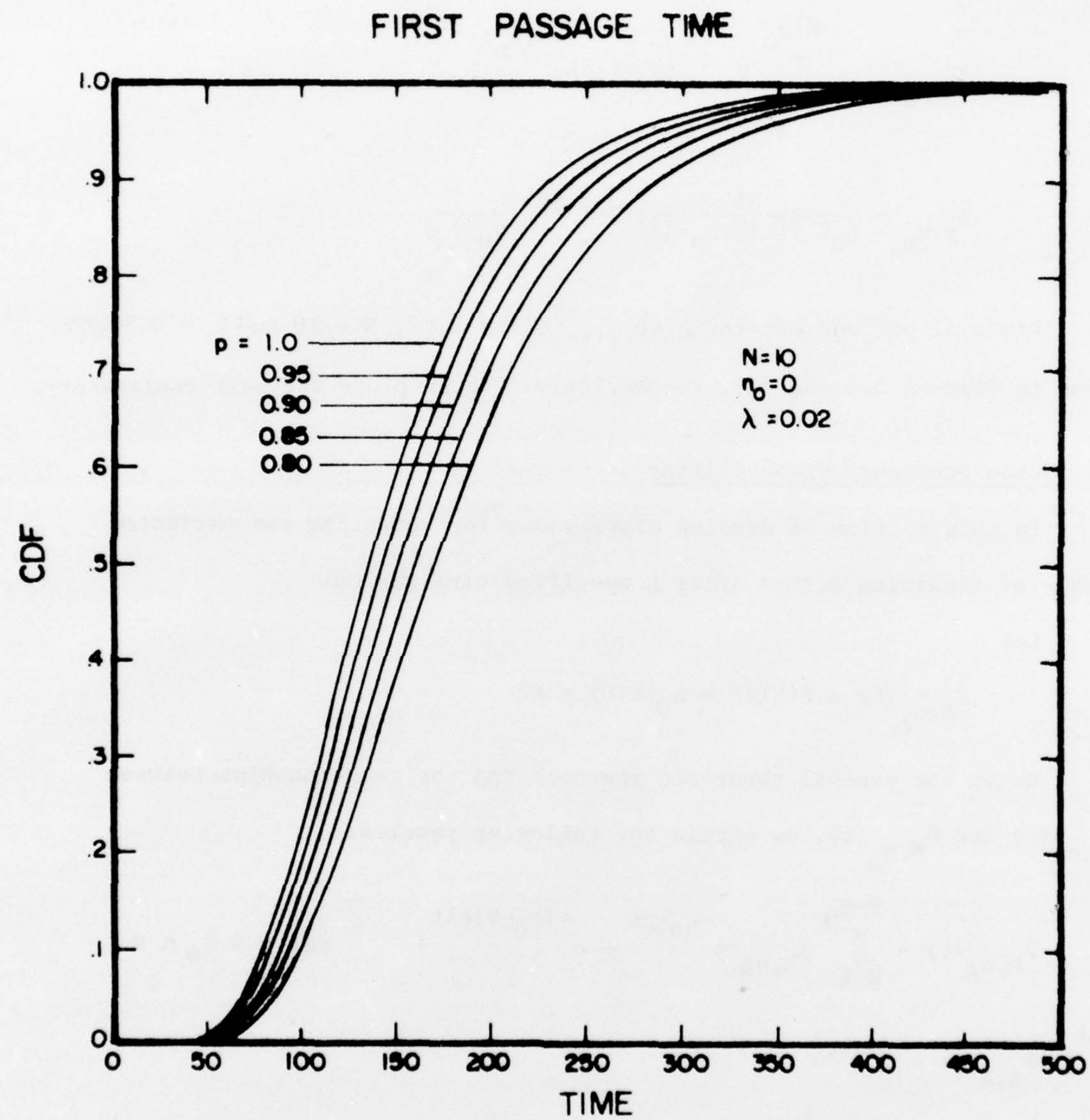


Figure 3.1. CDF of Time to Completely Debugged Software.

Using an approach similar to that of Section 3.3, we get the cdf and pdf as follows.

$$G_{N,n_0} = \sum_{j=1}^{N-n_0} B_{j,n_0}\{1 - e^{-(n_0+j)p\lambda t}\}$$

and

$$g_{N,n_0}(t) = \sum_{j=1}^{N-n_0} B_{j,n_0}(n_0+j)p\lambda e^{-(n_0+j)p\lambda t}$$

where

$$B_{j,n_0} = \frac{N!}{n_0!\ j!\ (N-n_0-j)!}(-1)^{j-1}\frac{j}{n_0+j}$$

Plots of pdf and cdf for $n_0=0,1,..,9, \lambda = 0.02$, $N = 10$ and $p = 0.9$ are given in Figures 3.2 and 3.3, respectively. These plots are self explanatory.

3.5 State Occupancy Probabilities

In this section we develop expressions for computing the estimated number of remaining errors after a specified time period.

Let

$$P_{N,n_0}(t) \equiv P\{X(t) = n_0 | X(0) = N\}$$

Using the renewal theoretic approach and the relationships between $P_{N,n_0}(t)$ and $G_{N,n_0}(t)$, we obtain the following results:

$$P_{N,n_0}(t) = \sum_{k=1}^{N-n_0} A_{k,n_0}\{e^{-n_0\lambda pt} - e^{-(n_0-k)p\lambda t}\} \qquad \text{for } 0 < n_0 < N$$

and

$$P_{N,N}(t) = e^{-N\lambda t}$$

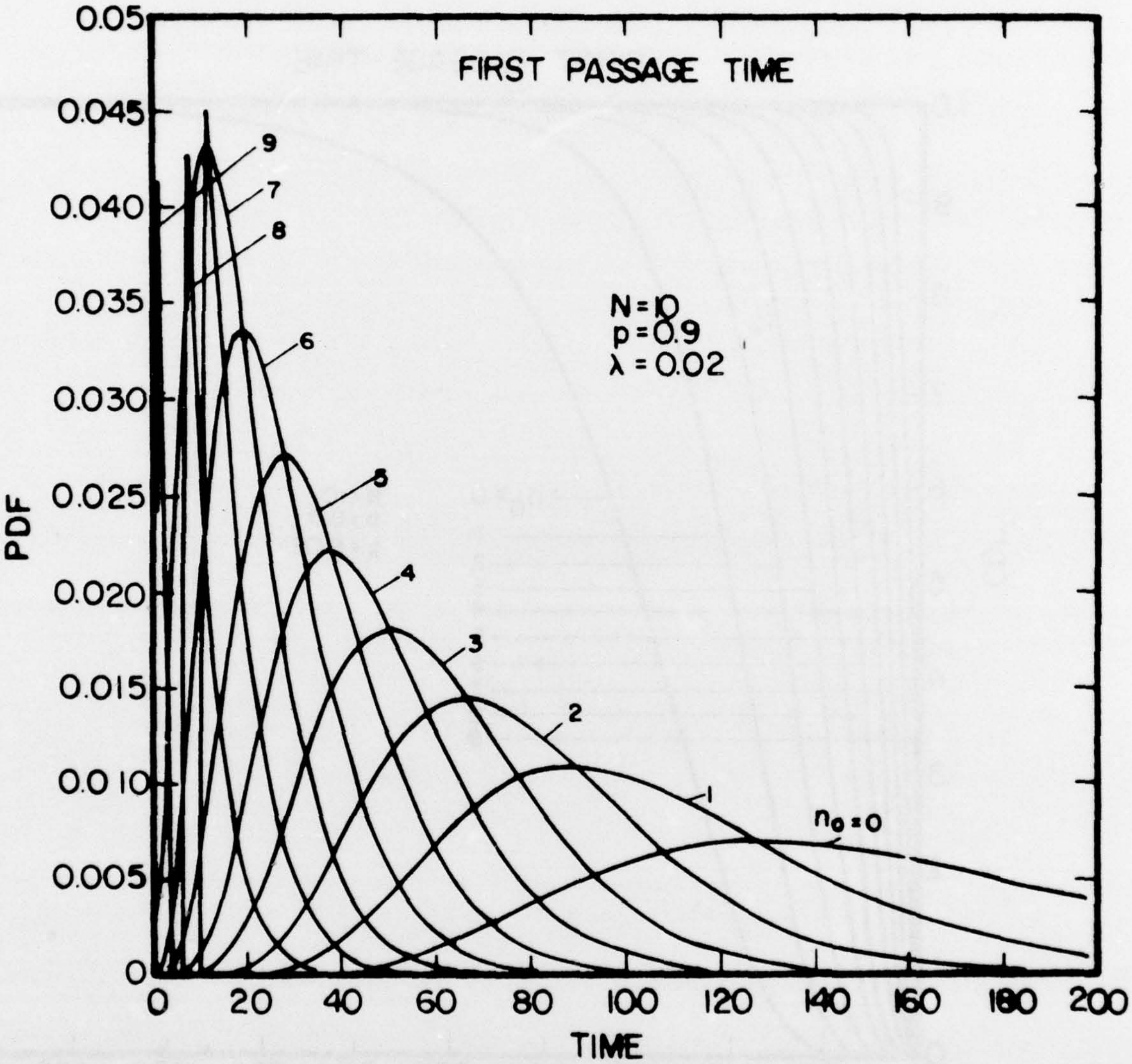


Figure 3.2. PDF of Time to a Specified Number (n_0) of Remaining Errors.

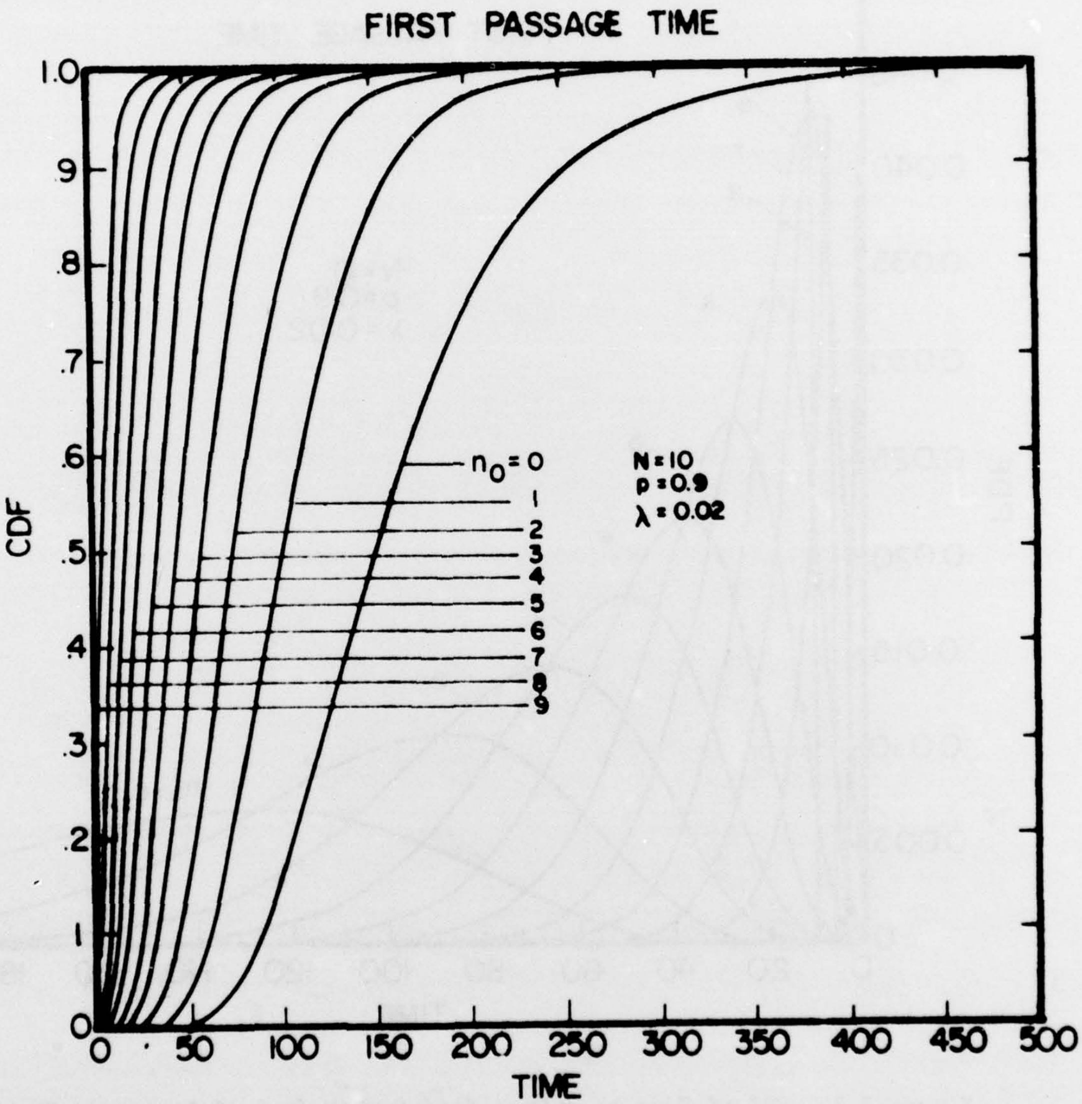


Figure 3.3. CDF of Time to a Specified Number (n_0) of Remaining Errors.

where

$$A_{k,n_0} = \frac{N!}{n_0!\ k!\ (N-n_0-k)!}$$

Further, an estimator of the number of errors remaining at time t is:

$$\hat{X}(t) = \sum_{n_0 \le N} n_0 P_{N,n_0}(t)$$

Plots of the probability distributions for $n_0 = 0,1,2,\ldots,10$ when $N = 10$, $p = 0.9$ and $\lambda = 0.02$ are given in Figure 3.4.

Now suppose we want an estimate of the number of errors remaining at time $t = 100$. We have

n_0	$P_{N,n_0}(100)$
10	0.152×10^{-7}
9	0.769×10^{-6}
8	0.175×10^{-4}
7	0.235×10^{-3}
6	0.208×10^{-2}
5	0.126×10^{-1}
4	0.530×10^{-1}
3	0.153
2	0.290
1	0.325
0	0.164

From these values we get

$$\hat{X}(100) = \sum_{n_0 \le 10} n_0 P_{N,n_0}(100) = 1.7$$

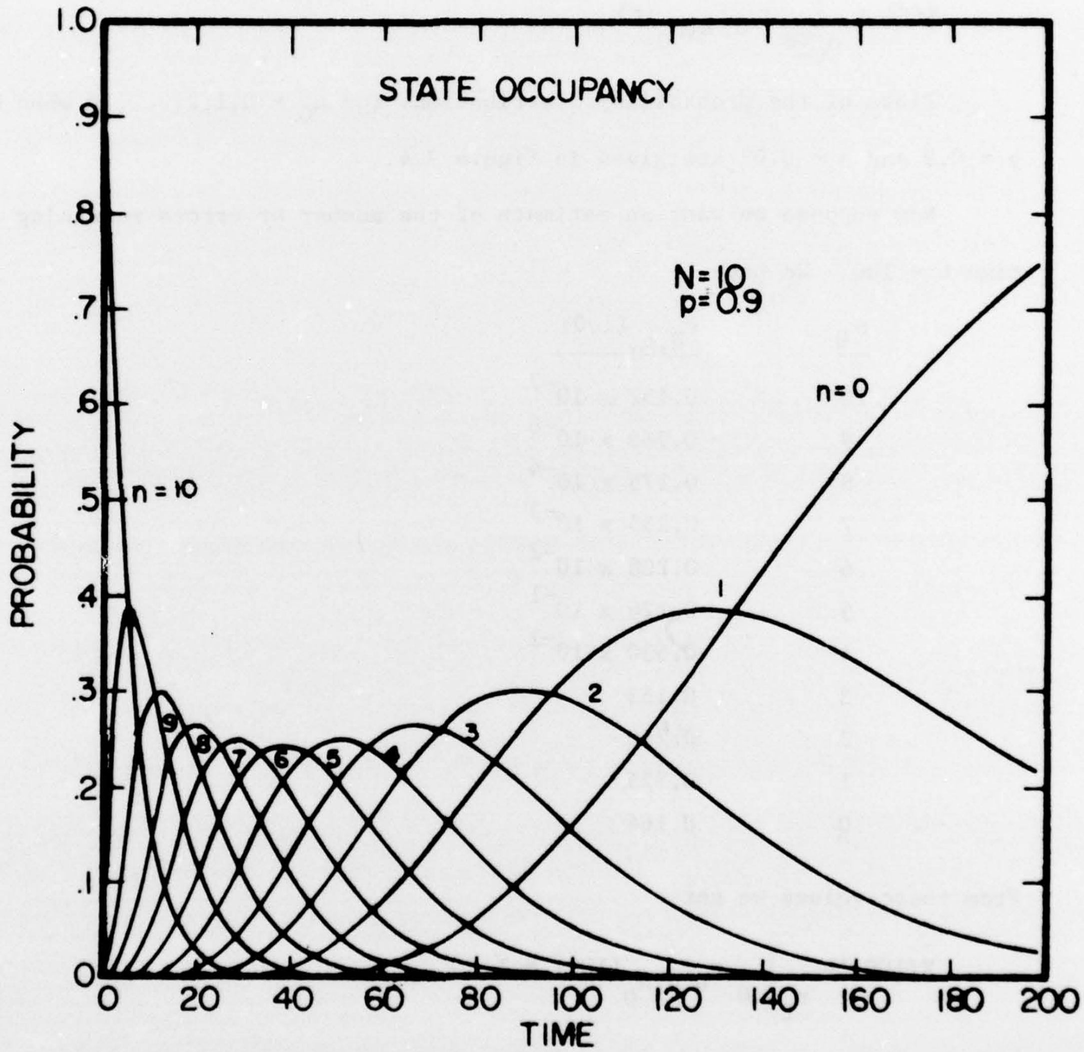


Figure 3.4. Probability Distribution for Given Number of Remaining Errors.

3.6 Expected Number of Errors Detected by Time t

Let N(t) be a random variable denoting the total number of errors detected by time t. Further, let $M_N(t)$ be the expected number of errors detected by time t when the initial number of errors is N, i.e.

$$M_N(t) = E[N(t)|X(0)=N]$$

Using Markov renewal theory and Laplace-Stieltjes transforms, we obtain the following results:

$$M_N(t) = \frac{1}{p} \sum_{k=1}^{N} \sum_{j=1}^{N-k+1} B_{j,k-1} \{1 - e^{-(k-1+j)p\lambda t}\}$$

Plots of $M_N(t)$ for N = 3,6,9,...30, λ = .02, and p = 0.9 are given in Figure 3.5 and are self explanatory.

3.7 Concluding Remarks

In the previous sections we described the key aspects of an imperfect debugging model. Expressions for various quantitites of interest were developed and numerical solutions were given for selected values of unknown parameters N, λ and p.

A detailed technical report on this task will be submitted to RADC in Spring 1977.

The model was developed on the assumption of zero correction time. For the follow-up model, we will include the time for corrective action. Also to be investigated is the problem of estimating the parameters from

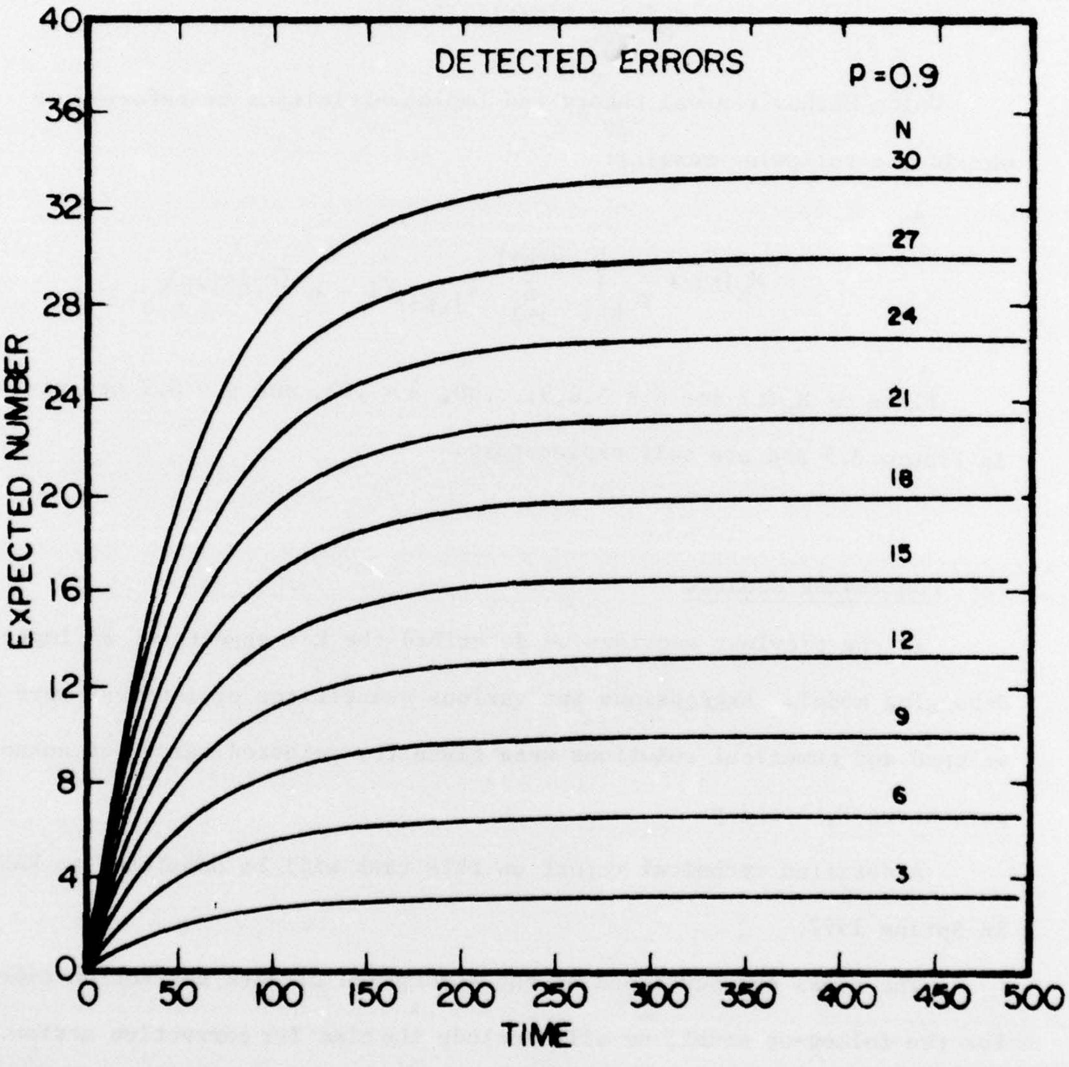


Figure 3.5, Mean Number of Detected Errors versus Time.

available data. In a Bayesian context, we plan to develop models when parameters N, λ and p have prior distributions.

Detailed technical reports on this task will be submitted to RADC as work is completed.

4. CURRENT WORK AND PLANS FOR THE NEXT PERIOD

The work on two modelling tasks was described in Sections 2 and 3. In this section we present an outline of the work being pursued and planned for completion in the next period.

Further Work on Software Reliability Modelling with Imperfect Debugging

The following aspects of this problem will be investigated:

(i) Development of a model to incorporate error correction time.

(ii) Estimation of parameters of the models being studied.

(iii) Development and study of a Bayesian imperfect debugging model

Expressions for various quantitites of interest will be developed and numerically studied.

Bivariate Software Error Model

We plan to develop software reliability models when the errors in the system can be classified as serious and non-serious. Various quantities of interest, for example system availability, will be studied.

Software Error Data Analysis

Some preliminary work is being pursued on the development of empirical models for software error data. We plan to continue this activity.

Demonstration and Testing Plans

Software reliability demonstration plans will be developed to provide a tool for making accept/reject decisions for software packages. Both classical and Bayesian plans will be investigated.

Sensitivity Studies

In this task we will investigate the effects of changes in prior distributions and/or model parameters on quantities of interest. The study will be mostly numerical in nature.

5. PERSONNEL

The following personnel have been involved in various aspects of the research described in Sections 2, 3 and 4.

Amrit L. Goel, Project Director and Principal Investigator

N. Bhalerao, Research Associate

K. Okumoto, Research Assistant

The research was also assisted on a short term basis by Messers R. Deb, C. Sastry, and S. Rao.

METRIC SYSTEM

BASE UNITS:

Quantity	Unit	SI Symbol	Formula
length	metre	m	...
mass	kilogram	kg	...
time	second	s	...
electric current	ampere	A	...
thermodynamic temperature	kelvin	K	...
amount of substance	mole	mol	...
luminous intensity	candela	cd	...

SUPPLEMENTARY UNITS:

plane angle	radian	rad	...
solid angle	steradian	sr	...

DERIVED UNITS:

Acceleration	metre per second squared	...	m/s
activity (of a radioactive source)	disintegration per second	...	(disintegration)/s
angular acceleration	radian per second squared	...	rad/s
angular velocity	radian per second	...	rad/s
area	square metre	...	m
density	kilogram per cubic metre	...	kg/m
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per metre	...	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm		V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	...	J/K
force	newton	N	kg·m/s
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m
luminance	candela per square metre	...	cd/m
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per metre	...	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m
magnetomotive force	ampere	A	...
power	watt	W	J/s
pressure	pascal	Pa	N/m
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	...	W/sr
specific heat	joule per kilogram-kelvin	...	J/kg·K
stress	pascal	Pa	N/m
thermal conductivity	watt per metre-kelvin	...	W/m·K
velocity	metre per second	...	m/s
viscosity, dynamic	pascal-second	...	Pa·s
viscosity, kinematic	square metre per second	...	m/s
voltage	volt	V	W/A
volume	cubic metre	...	m
wavenumber	reciprocal metre	...	(wave)/m
work	joule	J	N·m

SI PREFIXES:

Multiplication Factors	Prefix	SI Symbol
$1\ 000\ 000\ 000\ 000 = 10^{12}$	tera	T
$1\ 000\ 000\ 000 = 10^{9}$	giga	G
$1\ 000\ 000 = 10^{6}$	mega	M
$1\ 000 = 10^{3}$	kilo	k
$100 = 10^{2}$	hecto*	h
$10 = 10^{1}$	deka*	da
$0.1 = 10^{-1}$	deci*	d
$0.01 = 10^{-2}$	centi*	c
$0.001 = 10^{-3}$	milli	m
$0.000\ 001 = 10^{-6}$	micro	µ
$0.000\ 000\ 001 = 10^{-9}$	nano	n
$0.000\ 000\ 000\ 001 = 10^{-12}$	pico	p
$0.000\ 000\ 000\ 000\ 001 = 10^{-15}$	femto	f
$0.000\ 000\ 000\ 000\ 000\ 001 = 10^{-18}$	atto	a

* To be avoided where possible.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

AMERICAN REVOLUTION BICENTENNIAL 1776-1976